浅谈大数据技术在广电领域的应用

摘　要：21世纪是数据信息大爆炸的时代。随着网络新兴媒体、自媒体、微媒体的快速发展，广电传媒业遇到了前所未有的机遇与挑战。依托大数据、云计算、移动互联网等技术，广电传统媒体纷纷加快媒体融合的步伐。其中，通过大数据技术，可以面向互动电视端和移动客户端，通过分析用户行为数据进行个性化节目推送，为融媒体节目生产提供网络舆情热点线索，还可以解决广电领域面临的数据处理和存储的问题，本文从大数据概念出发，分析大数据技术，探讨大数据技术在广电领域的应用现状。

关键词：大数据技术；广电领域；应用
中图分类号：TP311.13　　文献标识码：A
文章编号：1671-0134（2018）04-059-03　　DOI：10.19483/j.cnki.11-4653/n.2018.04.020

文 / 高艳冬

在竞争日益激烈的全媒体时代下，要求广电传统媒体能在保证安全的前提下，与互联网进行多维融合发展，提供更丰富的资源、更多样的生产模式及更加快速权威的发布渠道。同时，面对海量的数据信息，对信息的处理和存储、采样分析、用户画像及可视化精准推送等功能，广电媒体面临着迫切需求，大数据技术的出现很好地解决了这个问题，推动了广电传媒的发展和转型。研究大数据技术、分析大数据技术在广电领域的应用，对于广电传媒而言，有着重要的意义。

1. 大数据技术

互联网催生了大数据应用的规模化环境，可以说，大数据技术是网络时代的衍生品，它是以计算机技术和信息技术为基础，实现对海量数据的收集、分析和处理，充分挖掘数据价值的一种技术。利用本地数据库资源、互联网的数据，实现数据源的提取，通常采用开源的分布式大数据技术架构，支持不同来源的异构数据的多种导入方式，实现海量数据的存储，通过对数据进行清洗、过滤、压缩等处理，实现大数据的预处理，利用数据索引功能和查询功能，可以实现数据库内的信息查询，利用云计算实现对海量数据的汇总和分析。在广电领域，大数据技术有着广泛的创新应用，主要有新闻舆情分析与智能选题、用户收视行为和效果分析、融媒体数据分析、产品精准推送、可视化展示等。

2. 大数据技术的构成

2.1 大数据平台构建技术

大数据平台构建技术包含了许多技术，并非单一存在，例如，分布式数据库、分布式文件系统、数据处理技术等，这些技术的应用，共同构成了大数据平台。

首先，分布式数据库技术。该技术主要是将网络技术同数据库技术相融合，主要形式有本地数据库、网络数据库等，利用网络技术，实现海量数据的存储。从技术层面进行分析，通过分布式数据库技术的实际应用，可以看出该项技术的核心在于CAP定理的使用，该技术保留了传统数据的一些特性，不同于传统数据库之处在于，可以实现对数据库信息的快速存取。当前，分布式数据平台的使用，以Hadoop、Greenplum、Hbase和MongoDB较为常见。

其次，分布式文件系统技术。一个优秀的分布式文件系统，可以随着数据规模的增加便捷地实现弹性横向扩展，部署大量的服务器支撑分布式文件系统技术的应用，这也就满足了海量数据的存储要求，提升了存储能力。为保证数据安全可靠使用，现今大多采用主/从双节点方式，当主名称节点工作时，从名称节点实时备份，与主名称节点数据同步，利用冗余存储的方式，可保证存储数据的安全性和可靠性。当前，分布式文件系统的使用，以Lustre、MogileFS、Hadoop、FreeNAS等较为常见。

最后，大数据处理技术。大数据处理技术应用十分广泛，在不同行业有着不同的表现形式，经过总结分析，可以将大数据处理技术归纳为以下三种形式：批处理和分析、实时流处理、近实时分析。

2.2 大数据挖掘技术

大数据技术的核心和关键在于对数据的挖掘，大数据挖掘技术涉及数据挖掘、分析和仓储三部分内容。

首先，大数据挖掘技术。大数据挖掘技术需要借助分布式计算平台来实现，以此为基础，实现对数据库海量数据的挖掘，为数据分析工作提供数据支持。大数据挖掘技术是将网络技术和传统的数据挖掘技术相结合，利用计算平台，对数据进行挖掘处理，遇到特殊情

况，可以进行简化操作，保证底层平台的有效运行。

其次，大数据分析技术。大数据分析技术是以大数据存储、管理和处理技术为基础，有目的地处理数据，实现数据的分析，将数据转换为报表。常见的数据分析技术有：日志分析技术、文件属性分析技术、文本情感分析技术、自然语言处理技术等。

最后，大数据仓库技术。大数据仓库技术是指将搜集到的海量数据进行整合和存储。

3. 广电大数据平台概述

3.1 广电大数据平台的应用要求

广电系统对于大数据平台的使用有如下要求：

首先，业务方面的要求。广电系统需要搜集的数据信息有：互联网、电视、新媒体等用户的行为数据、舆情数据等，将这些信息使用大数据平台进行搜集并整合。大数据平台在收集到用户行为数据之后，需要对这些数据进行挖掘和分析，调查用户的行为模式，了解用户的喜好，可以为广电系统提供新的运行模式。根据收集到的用户数据，可以创建视频模型，直观显示和分析收视情况。例如，根据不同的栏目和频道信息，了解实时收看人数，实现视频智能推荐、实时查询等功能。除此之外，广电大数据平台还应该满足如下要求：（1）提供实际案例的能力以及创建具有结构化数据处理能力的商业大数据平台，数据平台支持 B/S 模式，可以实现可视化的管理和开发；（2）根据 MapReduce 分布式数据计算框架和 HDFS 分布式数据存储机制，实现海量数据的告诉访问以及快速运算；（3）建立基于 MPP 架构和列存储的数据库，实现 PB 级的数据分析，建立 HHDFS 技术扩展与封装的平台，实现 PB 级的非结构化数据处理与分析。

其次，性能方面的要求。广电大数据平台的建设，需要实现数据的收集、整理、分析、查询等方面的要求，因此，大数据平台必须有灵活、可控、可扩容的机制，同时在性能方面还需要满足如下要求：（1）数据加载速度方面，数据加载速度不能小于 1TB/h；（2）数据处理速度方面，数据转换的处理速度不能小于 1TB/h；（3）复杂数据的统计计算方面，复杂指标的处理，统计和计算速度不能小于 15s。

3.2 广电大数据平台的建设内容

广电大数据平台的建设包含如下内容：一是大数据平台的建设；二是个性化门户的建设。其中，以个性化门户建设尤为重要，差异化策略是广电行业竞争的重要策略。因此，广电大数据门户建设，必须注重个性化和差异化。

首先，构建大数据平台。构建广电大数据平台时，需要考虑广电行业的发展，未来可能出现的业务转型需求，因此，大数据平台需要支持结构化数据，利用计算机架构，体现出大数据平台数据全面、技术先进、便捷开放的特点。广电大数据平台的建设内容有：研究平台的搭建、运维平台的构建、安装操作系统、平台部署、节点动态增减、MapReduce 作业管理、平台监控、参数调优、平台配置、挖掘工具、系统算法等。

其次，广电大数据个性化门户建设。个性化门户建设基于大数据平台的建设，利用大数据平台拥有的数据收集、整合、分析的能力，建立广电企业个性化门户。个性化门户的建设，对于用户数据的收集和分析十分重要，其中包括：用户的访问记录、用户聊天数据、论坛发帖等数据信息，将这些数据信息通过去格式化的方式进行整合、存储，使用行为、关系、语义、视觉等不同的算法工具，对用户数据进行处理，以此为基础建设个性化门户。大数据平台下，利用强大的云计算能力，可以将广电用户根据不同标准进行分类，以此为基础，实现内容的精准推荐，给予用户个性化的选择。通常情况下，个性化门户建设需要注意以下内容：（1）用户行为数据分析。广电大数据平台建设，主要在于利用大数据平台，收集用户数据，分析用户喜好和行为，用户行为数据主要有浏览业务数据、回看业务数据、点播业务数据、广告业务数据、直播频道业务数据、时移业务数据等；（2）点播、直播过程中的关联推荐。所谓关联推荐，是指根据大数据收集到的用户数据进行分析，分析用户的行为和喜好，建立用户喜好模型，根据用户喜好内容进行相关联内容的推荐，向用户推荐感兴趣的内容，提升用户的满意度，同时根据用户喜好，也可以提升广告的投放精准度；（3）互联网的用户分析。互联网时代下，互联网和移动媒体十分普及，人们利用移动网络、互联网络获取感兴趣的信息。因此，用户数据的收集也离不开互联网用户数据，深入挖掘用户喜好，从移动网络、广电网络、互联网络三网相融合，收集用户数据，为片源引进、广告投放、内容制作、收视率评估等工作内容提供数据支持。个性化推荐的出现，是大数据应用的最好体现，大数据时代，根据用户的实时数据和点击行为，通过数据分析可以获得用户的喜好，向用户推送感兴趣的内容。近年来，各大媒体门户网站或者移动媒体等都借助大数据的信息收集和分析能力，发觉用户喜好，推送相关节目，提升用户的满意度。

4. 大数据技术在广电领域的应用

4.1 在广电系统运维管理方面的应用

运维管理是大数据技术在广电系统的应用方面之一。由于广电领域涉及到许多非结构化的数据，这些数据的存在容易增加广电系统运维的难度。随着科技的不断发

展和进步，广电行业面临越来越多的数据信息，为了更好地对广电系统进行运行管理和维护管理，需要广电企业转型，加强对大数据技术的应用，实现数字化的管理模式。具体而言，以数字化模式为基础，建设大数据平台，进行节目制作、数据存储等内容。广电系统由多个平台构成，包括批量步数平台、批量安装操作、运维平台、作业管理平台等。大数据技术在广电系统中的应用，当系统出现故障时，可以及时定位发生故障的位置，运维管理工作效率得到提升。此外，大数据技术的应用，可实现广电系统运行情况的实时分析，可以及时了解系统的运行情况、负载情况，对系统问题做出科学预测，可以有效排除一些故障，做到早发现、早治理，为广电系统的正常运行提供可靠保障。

4.2 在广电业务运营管理方面的应用

广电行业对于业务运营的要求主要有：对电视、互联网、新媒体等数据信息进行收集、整合、分析，借助这些数据分析，对用户的喜好和行为进行预测，根据预测结果开展各项业务。大数据技术的应用，可以将这些数据信息在大数据平台整合分析，完成用户偏好的合理预测，建立视频模型。例如，根据用户浏览数据，可以模拟用户画像，对用户的年龄、喜好、内容等进行智能分类，根据不同分类给出智能推荐。此外，用户的数量、收视时长等信息，可以成为决策数据，为广电业务运营给出科学的指导。在实际运作过程中，使用 MapReduce 分布式计算框架，实现对用户数据的快速访问和高速运算。使用 MPP 架构数据库，实现对 PB 级结构化数据的处理。使用 HDFS 技术，实现对 PB 级非结构化数据的处理。可以看出，大数据技术，使广电系统可以更加精准地分析用户的喜好，做到有针对性地投放推荐内容，满足用户需求，提升用户满意率。

4.3 在广电用户数据管理方面的应用

建设个性化门户，需要做好用户数据的管理工作，加强对用户数据的利用，将用户数据进行整合、分析，对用户的偏好和行为做出科学预测。应用大数据技术，可以加强对广电用户数据的管理。首先，借助网络，可以实现同用户的沟通交流，获得用户数据。目前，广电企业对于用户数据的收集，以网络用户为单位，通过采样模式实现数据的采集和分析，可以使用收集到的有限数据，反映用户的收视情况。近年来，手机、平板的普及，移动设备的使用广泛，必然会带来一批移动用户，面对这种情况，广电企业应该加强互联网、移动网络、广电网络三网的互动和融合，实现数据信息的共享，使用大数据技术对用户的共性和个性进行挖掘，引导广电企业进行片源引入、收视率评估、内容制作等工作。大数据技术的应用，使许多收视数据被合理地收集和利用起来，加强了对用户数据的分析和利用，用户数据价值得到发挥，同时提升广电企业决策水平。

4.4 在新闻舆情分析与智能选题方面的应用

通过探针和爬虫手段，主动抓取主流新闻网站、论坛、博客、微博、微信、移动客户端等网络媒体的各种信息数据，通过对海量的数据的大数据分析，提取出隐含的巨大信息，提供关键性的市场数据，从而可实现舆情引导、舆情管控，支持新闻发现和传播效果分析，实现对互联网媒体的传播效果分析挖掘，还可对热点新闻推荐、热点微博线索、相似事件关联、热点变化趋势、辅助事前报道选题，支持用户参入分析（阅读 / 分享 / 评论等），用户关注度及倾向性分析挖掘等。

结语

在媒体融合大潮中，广电行业面临着一场划时代的创新革命，从媒体的内容形态、传播途径到发布接收方式都发生了颠覆性的变革，广电媒体应该充分利用大数据技术，研发建设适应广电应用的大数据技术平台，实现从 IT 时代向 DT 时代的融合转型。媒

参考文献

[1] 宋文娟 . 浅谈大数据技术及在广电系统中的应用 [J]. 辽宁广播电视技术，2014（4）：13-15.

[2] 王巨森 . 大数据技术在广电领域的应用探索 [J]. 网络安全技术与应用，2016（2）：9-10.

[3] 林健 . 大数据技术在广电领域的应用分析 [J]. 视听，2017（8）.

[4] 檀杉 . 大数据技术在广电领域的应用探索 [J]. 现代工业经济和信息化，2016，6（23）：79-80.

[5] 刘欣荣，陈凯 . 大数据技术在广电领域的应用探索 [J]. 工程技术：引文版，2016（6）：00291-00292.

[6] 李蕾 . “大数据，” 下广电行业的应用研究 [J]. 通讯世界，2017（13）：291.

（作者单位：吉林电视台）